

Ordinary Least Squares Estimation

January 20, 2025

1. Ordinary Least Squares Estimation

선형회귀분석은 독립변수 x 와 종속변수 y 사이의 선형적인 관계를 나타내는 수학적 모델이며, 독립변수의 개수에 따라 단순회귀분석(Simple Linear Regression)과 다중회귀분석(Multiple Linear Regression)으로 구분된다. 선형회귀분석 모델의 설명력을 극대화하기 위해서 오차(residual, ε) 제곱의 합을 최소화하는 회귀계수(parameter, β)의 값을 찾는 것이 최소제곱법(Ordinary Least Squares Estimation)이다.

2. Simple Linear Regression

단순선형회귀분석 모델은 다음과 같이 표현될 수 있다.

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, 2, \dots, n, \quad (1)$$

최소제곱법은 오차(잔차, residuals)의 제곱의 합(sum of squared residuals)을 최소화하는 회귀계수(β_0, β_1)의 값을 찾는 것이다. 잔차는 실제 값과 예측 값의 차이로 표현되며, 잔차 제곱의 합은 아래와 같이 정의될 수 있다. residuals:

$$e_i = Y_i - \hat{Y}_i \quad (2)$$

residuals sum of squares:

$$RSS = \sum_{i=1}^n e_i^2 \quad (3)$$

보다 자세히 나타내면 다음과 같다.

$$RSS = \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_i)^2 \quad (4)$$

Residual Sum of Squares를 최소화하기 위해서, β 에 대해 RSS를 미분한다. 우선 β_0 에 대해 미분하면 다음과 같다.

$$\frac{\partial RSS}{\partial \beta_0} = -2 \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_i) \quad (5)$$

다음으로, β_1 에 대해 미분하면 다음과 같다.

$$\frac{\partial RSS}{\partial \beta_1} = -2 \sum_{i=1}^n X_i (Y_i - \beta_0 - \beta_1 X_i) \quad (6)$$

위 미분 결과를 0으로 설정하면 최적의 β_0, β_1 을 찾을 수 있다.

$$\sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_i) = 0 \quad (7)$$

$$\sum_{i=1}^n X_i (Y_i - \beta_0 - \beta_1 X_i) = 0 \quad (8)$$

최종적인 해는 다음과 같이 구해진다.

$$\beta_0 = \bar{Y} - \beta_1 \bar{X} \quad (9)$$

$$\beta_1 = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2} \quad (10)$$

3. Multiple Linear Regression

다중선형회귀분석 모델은 다음과 같이 표현될 수 있다.

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \varepsilon_i, \quad i = 1, 2, \dots, n, \quad (11)$$

최소제곱법은 단순회귀분석과 마찬가지로 잔차 제곱의 합을 최소화하는 회귀계수들 $(\beta_0, \beta_1, \dots, \beta_k)$ 의 값을 찾는 것이다. 다중회귀분석에서의 잔차 제곱의 합은 다음과 같이 표현된다.

$$RSS = \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_{i1} - \beta_2 X_{i2} - \cdots - \beta_k X_{ik})^2 \quad (12)$$

$$RSS = \|\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\|^2 \quad (13)$$

RSS를 각각의 β_j ($j = 0, 1, \dots, k$)에 대해 편미분하면 다음과 같다.

$$\frac{\partial RSS}{\partial \beta_0} = -2 \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_{i1} - \beta_2 X_{i2} - \cdots - \beta_k X_{ik}) \quad (14)$$

$$\frac{\partial RSS}{\partial \beta_j} = -2 \sum_{i=1}^n X_{ij} (Y_i - \beta_0 - \beta_1 X_{i1} - \beta_2 X_{i2} - \cdots - \beta_k X_{ik}) \quad (15)$$

위 미분 결과를 0으로 설정하면 최적의 β 의 해는 다음과 같이 표현된다.

$$\boldsymbol{\beta} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y} \quad (16)$$